Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/36/20230461

Research on the model of automatic recognition and natural language question-answer system for traditional Chinese medicine tongue images based on LLMs

Yan Yang[1,2]**, Yunxia Yin**[1]**, Zhi Li**[1]

[1]School of Medical Information Engineering, Anhui University of Traditional Chinese Medicine, Hefei, Anhui, 230012, China

[2]yangyan@ahtcm.edu.cn

Abstract. Large Language Models (LLMs) have recently demonstrated their potential in clinical applications by providing valuable medical knowledge and recommendations. Traditional Chinese tongue diagnosis, one of the "Four Diagnoses," is an essential method for traditional Chinese medicine diagnosis. This paper builds upon tongue image classification technology and utilizes natural language processing and image recognition techniques to enhance the discrimination and analysis of traditional Chinese tongue images through learning and inference. We propose a method to integrate LLMs into the tongue image automatic recognition model and use them in interactive question-answering by summarizing and reorganizing information in natural language text format.

Keywords: Tongue Image Automatic Recognition, Natural Language Processing, LLM.

Funding: Anhui Provincial Quality Engineering Project of Higher Education Institutions: Medical Information Engineering Professional Transformation and Enhancement Project (2022zygzts056)

1. Introduction

In recent years, deep learning technology has shone brightly in the field of intelligent traditional Chinese tongue diagnosis. Deep learning automatically extracts high-dimensional features related to tongue images from a large amount of data, eliminating the need for complex feature engineering and generating a series of vectors. This project proposes a method to integrate the output of deep learning for traditional Chinese tongue diagnosis into LLMs. It leverages natural language dialogue systems to present information in natural language text format through summarization and reorganization.

Recently, Large Language Models (LLMs) have developed rapidly. Their ability to process and understand vast amounts of data makes them excellent in addressing complex issues. This research project combines the advantages of traditional Chinese medicine knowledge and logical reasoning with visual data by integrating LLMs. This advancement represents a significant step forward in the intelligent theory of traditional Chinese medicine. The paper employs stylized image captions to convert tongue image data into text. Compared to objective image captions, stylized image captions are a relatively new approach, designed to address the lack of style knowledge in image captioning technology. Modern natural language processing, like ChapGpt4, includes sentiment analysis, enabling responses to text in different styles. This paper explores the integration of LLM's sentiment analysis with earlier conclusions on tongue image constitution recognition to facilitate natural language responses.

2. Related Theories and Technologies

Automatic recognition of traditional Chinese tongue images has become a hot topic of research both domestically and internationally, while large language models (LLMs) represent a recent and cutting-edge research focus.

(1) Large Language Models: Large Language Models (LLMs) are advanced artificial intelligence systems that have been extensively trained on vast textual data [1,2]. These models employ deep learning techniques to generate human-like responses, making them suitable for a variety of tasks, including language translation, question-answering, and text generation. LLMs like OpenAI's GPT-3 [3] have made significant advancements in natural language processing. In the field of medicine, LLMs have shown potential as valuable tools for providing medical knowledge and recommendations. The Transformer architecture [4] and recent advancements in computational capabilities have enabled the training of large language models with billions of parameters, significantly enhancing their capabilities in summarization, translation, prediction, and generating human-like text [3,5,6]. Several domain-specific LLMs have been developed using pre-trained weights and training strategies. Examples include BioBERT [7] and PubMedBERT [8], which are trained on biomedical data from PubMed, and ClinicalBERT [2], which is further fine-tuned on the MIMIC dataset, outperforming its predecessor. Med-PaLM [6], developed towards the end of 2022 using curated biomedical corpora and human feedback, has shown promising results, including an accuracy rate of 67.6% on the MedQA exam. ChatGPT, without supplementary medical training, passed all three parts of the USMLE, achieving over 50% accuracy in all exams, with most exams exceeding 60% accuracy [9].

(2) Visual Language Models for Image Captioning: The renowned deep learning research institution, Google, introduced the "Deep Dream" computer vision project back in 2014. This project applied iterative and optimization algorithms within deep neural networks to enhance training effectiveness. It is evident that deep learning has found extensive applications in computer vision. A popular method for converting visual information into language is through image captioning. Image captioning models based on deep learning [10,11] can generate descriptive and coherent captions using large datasets like Microsoft COCO and Flickr 30K. In medical image analysis, image captioning methods are used to generate diagnostic image reports. For instance, Li et al. [12] achieved explicit learning of medical anomalies for report generation. Zhang et al. [13] utilized pre-built knowledge graphs based on disease topics. Another research direction [14] leverages self-attention architecture for cross-modal modeling. Recently, basic models with more clinical knowledge are expected to be a potential future direction. With an increase in model size, recent advancements in this field have shifted towards visual language pre-training (VLP) and utilizing pre-trained models. CLIP [15] merges visual and language information into a shared feature space, setting new state-of-the-art performance for various downstream tasks. Frozen [16] fine-tunes image encoders, with their outputs serving as soft prompts for language models.

(3) Computer-Aided Tongue Diagnosis: In the mid-1980s, the University of Science and Technology of China, in collaboration with Anhui University of Traditional Chinese Medicine, initiated the objective study of traditional Chinese tongue diagnosis using computer image processing and recognition techniques. Sun Liyou and others converted tongue images into digital images, combined different body information with symptoms for tongue analysis, and, with reference to different tongue diagnosis information, made diagnoses, completing exploratory experiments on quantitative analysis of tongue color discrimination. Subsequently, research institutions and universities such as the Beijing Institute of Chinese Medicine, Tianjin University of Traditional Chinese Medicine, and Beijing University of Technology embarked on research into computer-aided tongue diagnosis, making outstanding contributions in the fields of tongue chromatics, tongue segmentation, tongue image feature analysis, tongue image diagnosis, and the development of tongue image instruments.

In 2020, Song Haibei [17] and others combined machine learning technology with image processing methods to propose an AI-based diagnostic system for tongue and facial diagnosis. In recent years, related research in the field of traditional Chinese medicine has turned towards deep learning. Hu Jili and others [18] constructed a tongue image recognition and classification model on the TensorFlow platform, introducing convolutional neural network image processing techniques into the system for recognizing traditional Chinese constitution. This provides new research ideas and means for objective traditional Chinese constitution recognition, assisting doctors in rapidly diagnosing constitution and improving their work efficiency. The Lu Xun Academy Key Laboratory publicly published an article stating, "Deep learning can automatically extract features for precise identification of tongue coating, tongue body, tongue edges, tongue roots, and other tongue characteristics. The recognition accuracy for tongue coating images exceeds 95%, and the accuracy for tongue fissure images surpasses 90%." Furthermore, some relevant research achievements have already been made. For example, Baidu AI introduced a system called "Tongue Diagnosis AI" in 2019. This system automates traditional Chinese tongue diagnosis through the combination of deep learning and big data analysis, enabling the rapid and accurate processing and analysis of human tongue diagnosis data.

3. Research Methodology

(1) Establishment of Automated Tongue Image Analysis Model

Currently, the proposed scheme for the automation of tongue diagnosis analysis process is shown in Figure 1. Initially, tongue images are collected and subjected to image preprocessing, including color correction. Subsequently, the tongue body is segmented and various visual features of the tongue body are recognized and extracted. Finally, relevant knowledge rules from the traditional Chinese medicine diagnosis knowledge base are utilized to infer from tongue image features to Chinese medical syndromes.

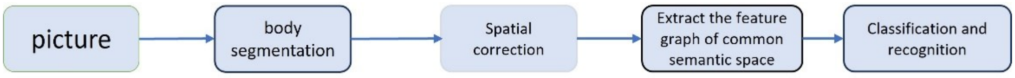

picture
body segmentation
Spatial correction
Extract the feature graph of common semantic space
Classification and recognition

Figure 1. Basic Architecture of the Tongue Image Feature Recognition Model.

(2) Output in Natural Language Text

The first part of the tongue image analysis model, which will be developed using deep learning algorithms, is combined with natural language processing and image recognition techniques. Through learning and inference, it aims to enhance the discrimination and analysis levels of traditional Chinese tongue images and facilitate applications such as tongue image classification output.

(3) Human-Machine Dialogue Model

Based on the aforementioned results and language models trained on medical knowledge, the model engages in dialogues related to symptoms, diagnosis, and treatment. It provides analysis and outputs to patients based on symptoms, diagnosis, treatment methods, dietary advice, and other information.

4. Experimental Models

4.1. Tongue Image Classification Model Experiment

(1) Data Source and Preprocessing

The tongue image data in this study originate from an affiliated hospital of a certain traditional Chinese medicine university. Images have been labeled based on patients' tongue diagnosis information. A portion of the image dataset is selected as training data. Each category contains two thousand images. Since the original tongue image resolution is 2592x1728px, which is excessively large and would consume significant memory resources during training, the original images are resized to a resolution of 600px using the Python library PIL.

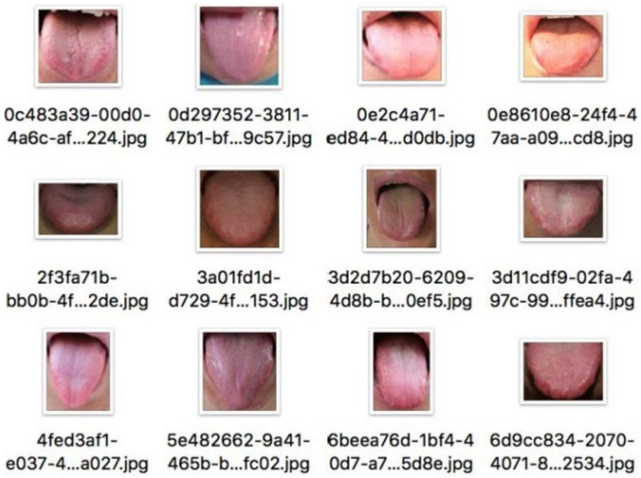

0c483a39-00d0-4a6c-af...224.jpg
0d297352-3811-47b1-bf...9c57.jpg
0e2c4a71-ed84-4...d0db.jpg
0e8610e8-24f4-47aa-a09...cd8.jpg
2f3fa71b-bb0b-4f...2de.jpg
3a01fd1d-d729-4f...153.jpg
3d2d7b20-6209-4d8b-b...0ef5.jpg
3d11cdf9-02fa-497c-99...ffea4.jpg
4fed3af1-e037-4...a027.jpg
5e482662-9a41-465b-b...fc02.jpg
6beea76d-1bf4-40d7-a7...5d8e.jpg
6d9cc834-2070-4071-8...2534.jpg

Figure 2. Tongue Image Dataset.

(2) Experimental Process

The experimental process is depicted in Figure 3.

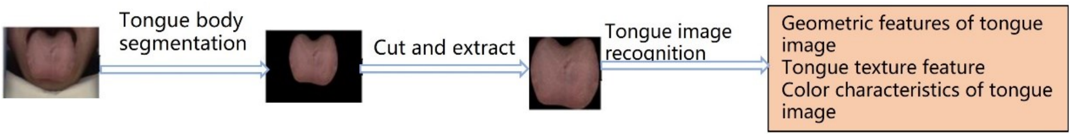


Figure 3. Experimental Process for Tongue Image Classification Model.

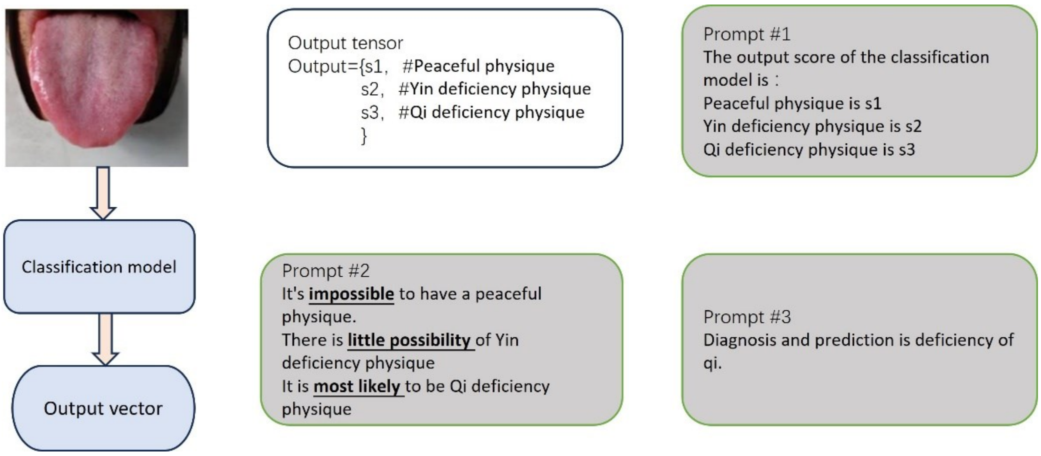
The first part involves tongue body segmentation. Deep learning and transfer learning techniques will be used to extract geometric features, texture features, and color features from tongue images.

The second part focuses on tongue feature recognition, which includes geometric and texture feature recognition for the tongue body (tongue tissue and tongue coating). A transfer learning-based tongue image analysis will be proposed. It employs a combination of geometric feature analysis methods involving STN and VGG16, as well as texture feature analysis methods involving LREL and VGG16.

4.2. Large Language Model Training Experiment

Since GPT-3 is open source, we plan to use OpenAI's publicly accessible API, which provides four different sizes of GPT-3 models: text-ada001, text-babbage-001, text-curie-001, and text-davinci003.

The first part of the tongue image analysis model, developed using deep learning algorithms, will be trained. It will be combined with natural language processing and image recognition techniques to improve the discrimination and analysis of traditional Chinese tongue images through learning and inference, with the aim of achieving applications such as tongue image classification output.



Output tensor
Output={s1, #Peaceful physique
s2, #Yin deficiency physique
s3, #Qi deficiency physique
}

Prompt #1
The output score of the classification model is：
Peaceful physique is s1
Yin deficiency physique is s2
Qi deficiency physique is s3

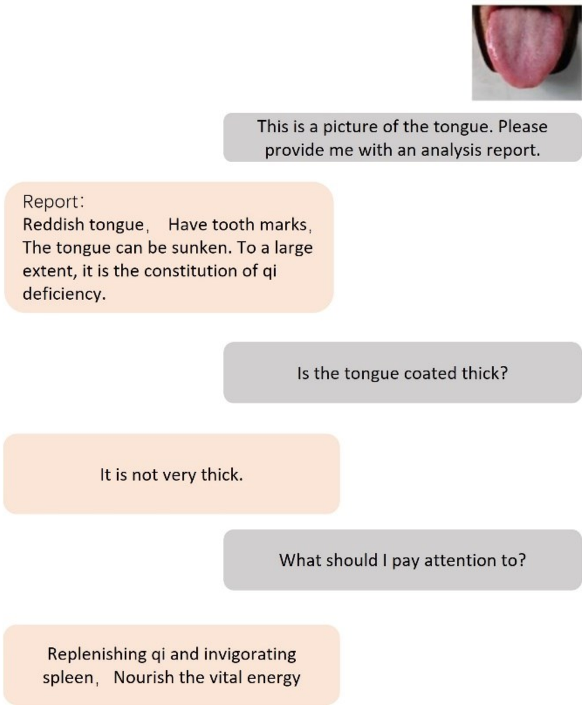
Prompt #2
It's **impossible** to have a peaceful physique.
There is **little possibility** of Yin deficiency physique
It is **most likely** to be Qi deficiency physique

Prompt #3
Diagnosis and prediction is deficiency of qi.

Figure 4. Conversion of Classification Model to Text.

4.3. *Natural Language Dialogue System*

Based on the results obtained and language models trained on medical knowledge, the system engages in dialogues related to symptoms, diagnosis, and treatment, as shown in Figure 5. It provides analysis and outputs to patients based on symptoms, diagnosis, treatment methods, dietary advice, and other information.



This is a picture of the tongue. Please provide me with an analysis report.

Report:
Reddish tongue, Have tooth marks, The tongue can be sunken. To a large extent, it is the constitution of qi deficiency.

Is the tongue coated thick?

It is not very thick.

What should I pay attention to?

Replenishing qi and invigorating spleen, Nourish the vital energy

Figure 5. Interactive Diagnosis with ChatGpt4.

5. Conclusion

This research conducted experiments using modern artificial intelligence technology to develop a large language model based on traditional Chinese tongue images. This model can rapidly generate natural language descriptions for various indicators and parameters of tongue images. It not only reduces the workload of traditional Chinese medicine diagnosticians but also enhances the accuracy and standardization of traditional Chinese tongue diagnosis. This represents a significant breakthrough in the field of artificial intelligence in traditional Chinese medicine. By employing deep learning technology and combining it with tongue image representation learning, the study analyzed and extracted critical features from tongue image data, leading to a more accurate and comprehensive model for assisting in traditional Chinese tongue diagnosis. This approach, which leverages deep learning technology, offers a wide range of applications in the field of traditional Chinese tongue image-assisted diagnosis.

The close integration of traditional Chinese medicine and artificial intelligence explores the intersection of research between traditional Chinese medicine and artificial intelligence, providing new insights for the future of intelligent healthcare, health examinations, and related fields. It lays a solid theoretical and practical foundation for interdisciplinary research between traditional Chinese medicine and artificial intelligence.

The use of large language models, in combination with medical imaging and natural language, facilitates effective communication and interaction between doctors and patients, ultimately improving the accuracy and standardization of traditional Chinese tongue diagnosis. By incorporating domain knowledge from both traditional Chinese medicine and artificial intelligence, this approach offers new ideas and methods for cross-application between these two fields, fostering the development and integration of both domains and significantly advancing the informatization of traditional Chinese medicine.

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprintarXiv:1810.04805, 2018. 1, 3

[2] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. Improving language understanding by generative pre-training. 2018. 1

[3] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. Advances in neural information processing systems, 33:1877–1901, 2020. 1, 3

[4] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in neural information processing systems, pages 5998–6008, 2017. 3

[5] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee,Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. Exploring the limits of transfer learning with a unified text-to-text transformer. The Journal of Machine Learning Research, 21(1):5485–5551, 2020. 3

[6] Karan Singhal, Shekoofeh Azizi, Tao Tu, S Sara Mahdavi,Jason Wei, Hyung Won Chung, Nathan Scales, Ajay Tanwani, Heather Cole-Lewis, Stephen Pfohl, et al. Large language models encode clinical knowledge. arXiv preprint arXiv:2212.13138, 2022. 3

[7] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, DonghyeonKim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. Biobert:a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4):1234–1240,2020. 3

[8] Yu Gu, Robert Tinn, Hao Cheng, Michael Lucas, Naoto Usuyama, Xiaodong Liu, Tristan Naumann, Jianfeng Gao,and Hoifung Poon. Domain-specific language model pretraining for biomedical natural language processing. ACM Transactions on Computing for Healthcare (HEALTH),3(1):1–23, 2021. 3

[9] Tiffany HKung, Morgan Cheatham, Arielle Medinilla, ChatGPT, Czarina Sillos, Lorie De Leon, Camille Elepano, Marie Madriaga, Rimel Aggabao, Giezel Diaz-Candido, et al. Per-formance of chatgpt on usmle: Potential for ai-assisted medical education using large language models. medRxiv, pages 2022–12, 2022. 3

[10] Simao Herdade, Armin Kappeler, Kofi Boakye, and Joao Soares. Image captioning: Transforming objects into words.Advances in neural information processing systems, 32,

2019. 3

[11] Quanzeng You, Hailin Jin, Zhaowen Wang, Chen Fang, and Jiebo Luo. Image captioning with semantic attention. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 4651–4659, 2016. 3

[12] Guanxiong Liu, Tzu-Ming Harry Hsu, Matthew McDermott,Willie Boag, Wei-Hung Weng, Peter Szolovits, and Marzyeh Ghassemi. Clinically accurate chest x-ray report generation.In Machine Learning for Healthcare Conference, pages 249269. PMLR, 2019. 3

[13] Yixiao Zhang, Xiaosong Wang, Ziyue Xu, Qihang Yu, Alan Yuille, and Daguang Xu. When radiology report generation meets knowledge graph. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, pages 12910–12917, 2020. 3

[14] Zhihong Chen, Yaling Shen, Yan Song, and Xiang Wan.Generating radiology reports via memory-driven transformer. In Proceedings of the Joint Conference of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Aug. 2021. 3, 4, 6

[15] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry,Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In International Conference on Machine Learning,pages 8748–8763. PMLR, 2021. 3

[16] Maria Tsimpoukelli, Jacob L Menick, Serkan Cabi, SM Eslami, Oriol Vinyals, and Felix Hill. Multimodal few-shot learning with frozen language models. Advances in Neural Information Processing Systems, 34:200–212, 2021. 3

[17] Song, H., Wen, C., & Cheng, X. (2020). Construction of an AI-based Traditional Chinese Medicine Tongue and Facial Diagnosis Assistance System. Shi Zhen National Medicine and Traditional Chinese Medicine, 31(02), 502-505.

[18] Hu, J., & Kan, H. (2018). Tongue Image Classification Based on Convolutional Neural Networks. Journal of Anqing Normal University (Natural Science Edition), 24(04), 44-49. https://doi.org/10.13757/j.cnki.cn34-1328/n.2018.04.010.